机器人打开数据传送门，与你互动跑两会
——访 i 思工作室负责人李杰

本刊记者 | 陈旭管



李　杰

i 思，出生于 2017 年 2 月 19 日新华社音视频部 i 思工作室，3 月两会期间正式成为新华社见习记者，参与两会报道，既可以采访代表委员，搜集百姓关心的两会话题，也可以在演播室与主持人、嘉宾互动，与此同时还具备两会大数据分析、查询代表委员信息、查询天气、两会知识问答等功能。据了解，新华社客户端《i 思跑两会》系列与提及 i 思机器人的短视频报道，累计浏览量达 4470 多万人次。

本刊记者，走近新华社音视频部对话 i 思工作室负责人李杰，畅谈人工智能背后两会内容报道创新的路径探寻。

i 思档案：

昵称：i 思

英文名：Inspire

年龄：不到一岁

性别：我是小男生

身高：1.28 米

体重：60gk

生日：2 月 19 日

出生地：北京 新华社音视频部 i 思工作室

职业：新华社见习记者

语言：普通话、一点点英语

小伙伴：人民日报小融

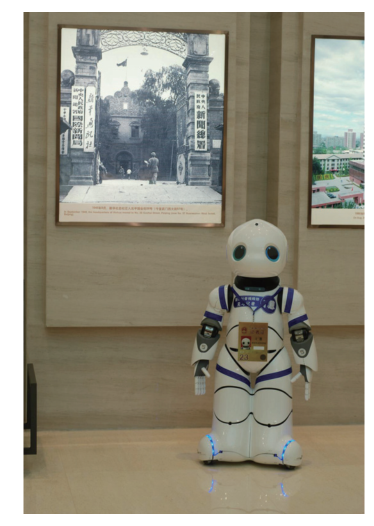
梦想：两会上向总理提问

《中国传媒科技》：两会期间，i 思作为新华社见习记者参与两会报道，并进行视频直播节目的报道，最初的设计是什么？

李杰：其实从 2016 年 9 月份，我们就萌生过创造机器人记者的想法。当时，《新华视点》栏目在音视频部领导的带领下，正在做有关人工智能的重点调研项目，专题室编导一边学习人工智能知识，一边组织策划、采访编辑。为了做好调研选题，我们专门建立了人工智能微信群，交流有关前沿技术发展趋势，还请了人工智能学科专家来做讲座，用了 3 个月时间进行人工智能方面的专业学习。到 2016 年 12 月，完成了《2016 人工智能元年的中国机遇系列》6 集系列专题片，作为国内首个全面报道人工智能领域题材的新闻专题，还获得了社级好稿的评价。在调研中和很多技术团队接触时，我们就发觉人工智能是媒体未来的发展方向。

2017 年 1 月筹备两会报道时，音视频部领导大力鼓励我们创新，于是我们就提出了围绕两会做一个新闻互动机器人的设想。我们最初的设想是做一个虚拟机器人，可以在手机上和受众互动两会新闻和有关知识。后来我们又提出了更高的目标——做新华社首个实体机器人，可以直接在视频节目中出镜、采访。于是就有了 i 思这样的实体机器人，i 思的出现不仅出色完成了两会内容报道的任务，而且还成为刷爆朋友圈的现象级产品，获得很多年轻受众的关注。

《中国传媒科技》：可否简单介绍 i 思工作室及技术研发情况？



李杰：i 思出生在今年 2 月 19 日，就是习总书记来新华社“2 · 19 讲话”一周年的日子。为落实好“2 · 19 讲话”，在两会创新报道中抢占先机，音视频部领导牵头策划指导，以视频专题编辑室为主体加上部里一些创意人才，成立了 i 思工作室。工作室联合社外技术团队，以二次创业精神投入到机

器人研发工作。i 思从大脑到硬件集成了中国当前人工智能前沿科技成果，大脑和硬件分别与海知智能和北京康力优蓝两家国内优秀的技术团队合作。

我们和技术团队对 i 思机器人整体角色、性格和语言风格以及知识体系进行了全方位的准备。大脑方面，基础和通用的知识根据新华社见习记者这个定位进行筛选、过滤和调整，形成符合新华社定位的知识图谱；两会知识图谱则是从零开始建立起来，以及对应的功能逻辑也进行了相应的设计实现。

硬件实体方面，我们和技术团队做了全面的测试以及多层次的调试，以及网络适应性调整等来保障现场效果。特别是对 i 思的唤醒方式、反应速度进行优化。因为软硬件是两个技术团队合作，前期需要大量的磨合和准备。

最终 i 思在外型硬件上以大眼睛、小身材、简单的动作设计为主，在语言风格上更加年轻态。将新华社气质与软萌的人物形象相结合，力求在两会报道中形成一种严肃活泼的报道风格。例如你问他：“两会放假吗？”i 思会用软萌的语气说：“刚过完年，又想放假，撸起袖子加油干吧！”

《中国传媒科技》：作为两会报道的见习记者，初次尝试重大主题的报道，其大数据选取的质量如何保障？

李杰：数据是供给机器人大脑的营养，需要产生机器人能够理解和掌握的知识。为了保证重大报道的安全、准确，i 思的大数据调取来源于全国人大、政协官方网站，以及新华网、中国政府网等网站，保证调取数据的准确性。比如，全国人大代表数据资源就是调取了全国人大官方网站最新更新的数据。另外对于更宽泛来源的数据，可通过数据清洗、融合和交叉计算等算法机制来保障有足够质量标准的知识成为机器人的技能。此外，i 思还有很多外援，比如新浪微博就特别提供了两会热词大数据分析，经过编辑梳理后提供给 i 思发布。

《中国传媒科技》：i 思大脑中目前包含哪些语料库，其规模如何？

李杰：人的大脑是一个复杂的机制，机器人大脑同样也不简单。针对两会和新华社记者应知信息我们已经逐渐完善形成了一个单独语料库。其中全国两会代表委员信息就有五千多条，对应语料库大概是十倍、百倍甚至更多的原始语料库加工处理而来。比如，可以问某位代表是谁，也可以直接问某位代表的家乡。

另外，海智智能 bot 平台还给 i 思提供了很多基础语料库，比如地点、百科等常识类词条有千万量级，古诗词 30 万首、成语 9 万首，覆盖国内区县和全球主要城市的实时天气，还有大量人性化的对话闲聊等。说一句古诗，i 思就能告诉你诗词的出处。天气方面，甚至我们可以问 i 思“明天会不会下雨”。

《中国传媒科技》：i 思在两会期间主要配合记者完成哪几方面的工作，在互动性上都可以满足哪些互动？

李杰：i 思配合记者编辑完成了 7 期《新华视点》《新华纵横》专题节目，共 90 分钟节目，i 思机器人作为采访记者、主持人出现在节目中。他比较擅长提供两会大数据分析，代表网友向嘉宾提问，还擅长将一些单调的问题转换得更通俗易懂，萌言萌语很受观众喜爱。黑龙江卫视特别开辟《i 思跑两会》专栏，将 i 思机器人作为特约记者，每天与电视台主持人进行连线，报道两会热点。



互动方面可以涉及代表委员资料查询、民意搜集、两会新闻推荐、两会知识问答、天气查询、古诗词互动、大数据分析等。

《中国传媒科技》：两会期间，接受访问的代表们反响如何，参与报道的视频内容收视率怎样？

李杰：i 思采访了不少代表和专家，比如采访来自四川大凉山的全国人大代表潘成英，关心悬崖村天梯小学的孩子。潘成英代表第一次见到这样智能的机器人记者特别惊讶，并热情邀请 i 思去大凉山采访，希望更多像 i 思这样的机器人能够支援贫困地区儿童教育。全国人大代表、中科院副院长张涛认为“新华社推出 i 思机器人记者非常生动活泼。特别是和年轻人互动得非常好，有利于把两会消息及时传达给社会。人工智能将带来下一代技术革命，机器人、人工智能将是一个非常重要的方向。”



法律专家龙卫球院长在做解读民法节目中也对 i 思大为赞赏，很多抽象的法律问题通过 i 思的提问更加通俗、生动、有趣。比如 i 思问法律专家“i 思今年 1 岁，等我到 6 岁时会有什么民事权利呀？”“以后像我这样的机器人越来越多，民法什么时候加入机器人的权利啊？”又引发了专家对这一讨论的极大兴趣。



科幻作家韩松老师在和i思一起录制访谈节目后说，"感觉特别科幻。"i思还提出了一个小目标，希望成为韩松老师科幻小说里的人物，韩松老师很开心地回答，"我一定会写一部小说，让你当主人公。"



《i思跑两会》系列短视频很多单条点击量都达到400多万人次。从用户年龄上看，其实不仅是小朋友和年轻人特别喜欢，中年人以及老年人也很关注。

《中国传媒科技》：据了解，i思做了一期英文海媒视频直播，外媒反响如何？

李杰：是的，这期视频直播节目在Twitter、Youtube、Face Book、Instagram海外四大平台播出。仅在Twitter上十几分钟直播就有68331人围观，还被大V推荐。很多网友留言说"这小机器人真可爱。""中国将会变得越来越强大！""让人印象非常深刻！"

其实在两会期间，i思机器人记者成为海外关注的热点，成为中国人工智能产业优秀成果的展示，本身就是一次成功的主动议题设置。今年，i思机器人记者在两会上亮相，恰逢人工智能首次写入政府工作报告，中国人工智能又处于行业发展的风口。从整体舆论导向上，i思不仅展示了新华社在智媒领域的创新和探索，也为中国的人工智能产业以及中国科技创新发展营造了良好的舆论氛围。

《中国传媒科技》：您作为产品研发成员之一，认为i思最大的特色、创新点在哪里？

李杰：i思作为首款实现两会采访的实体智能机器人，最大的特色是类似或者一定程度上接近于人的方式进行实际采访。

从创新点上来说，首先实现了软硬件一体化解决方案，语音识别、语义理解、语音合成以及部分动作表情等，能够顺畅地和采访对象对话。并且i思可以通过人脸识别和语义理解相结合与人交互，就是认出来人之后，了解对方的背景后再进行沟通和提问。

其次，i思不仅拥有两会以及相关采访知识，还掌握了上亿中文知识图谱，涵盖社会、文化、生活、天气等百科知识，并能动态学习更新。

与此同时，i思也是有效可控的机器人，是人工+智能机制的产物。比如从知识体系到机器学习训练机制上，都能开放给我们编辑和知识领域专家来操控，并不需要懂得太复杂的人工智能技术，只需要经过简单培训，掌握一套机器人大脑管理工具即可。

上述这些特色和创新点，其实是多种人工智能最新技术的集成，体现我国人工智能产业发展的最新成果。比如自然语言理解，通过词法、句法乃至篇章层面的机制，以及多层次机器学习手段来实现。比如通过深度学习里面的RNN（循环和递归神经网络）来做句子主干抽取，CNN（卷积神经网络）来做意图领域分类等；还有就是知识图谱技术的应用，从原始的各种类型和多源的数据，尤其是非结构化文本爬取、清洗、融合、交叉计算等机制，形成覆盖全面、更新及时以及有质量保障的知识图谱，实现i思理解、推理、推荐和记忆等问答交互功能。

《中国传媒科技》：在视频节目中，i思全程自己完成采访任务，那么其中的问题设计是如何实现的？

李杰：一方面是i思通过从大量的文本数据自动化提取观点或者问题的线索。一方面也支持人工直接干预，编辑可以设置提问的关键词，i思识别关键词就可以按照编辑意图回答了。我们的视频节目很多都原汁原味地展示了人机交互过程，细心的观众会发现，i思在提问前，主持人会对i思有些关键词提示，来引导i思向嘉宾提问。

《中国传媒科技》：机器人记者是否会成为未来重大主题报道的新趋势？

李杰：我想一定是媒体未来发展的方向，未来媒体比拼的将是人机协同的报道模式，机器人作为记者编辑的助手和工具，提高报道效率，也丰富报道内容，特别是在大数据分析方面有着先天的优势。所以我们要尽快去学习和利用人工智能技术，不能忽视或排斥它的存在。就像韩松老师在接受i思采访时说的那样，人工智能技术就像疾驰而过的列车，如果我们现在不登上这列列车，可能就被远远落在后面，再也追赶不上了。

《中国传媒科技》：在新华社音视频部内部采用工作室的形式参与产品创新，与本职工作之间如何协调，两会结束后，对这种内部成立工作室参与产品创新的工作模式有哪些思考？

李杰：i思工作室是在音视频部领导支持下创新机制成立的，成员都是兼职编辑，在完成本职工作基础上，以二次创业的精神自愿参加的。为了兼顾本职工作，我们用i思机器人完成了一些我们专题室的日常报道。

两会结束后，我们i思工作室还在尝试新的探索，未来交互式新闻模式一定是发展趋势，i思也会新增更多本领。我们希望有更多的媒体人投入到创新的大潮中来。媒